Programmed cell death-index (PCDi) as a prognostic biomarker and predictor of drug sensitivity in cervical cancer: a machine learning-based analysis of mRNA signatures

Wei Wang†, Pengchen Chen†, Songhua Yuan‡

1. Department of Obstetrics and Gynecology, The First People’s Hospital of Foshan, Foshan, 528000, Guangdong, China.
2. Dongguan Maternal and Child Health Care Hospital, Postdoctoral Innovation Practice Base of Southern Medical University, Gongguan, 523125, Guangdong, China.

† These authors contributed equally: Wei Wang, Pengchen Chen.
‡ Corresponding author: songhua.yuan1974@gmail.com.

© The author(s). This is an open access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/).

Abstract

Purpose: Cervical cancer is a significant public health concern, particularly in developing countries. Despite available treatment strategies, the prognosis for patients with locally advanced cervical cancer and beyond remains poor. Therefore, an accurate prediction model that can reliably forecast prognosis is essential in clinical setting. Programmed cell death (PCD) mechanisms are diverse and play a critical role in tumor growth, survival, and metastasis, making PCD a potential reliable prognostic marker for cervical cancer.

Methods: In this study, we created a novel prognostic indicator, programmed cell death-index (PCDi), based on a 10-fold cross-validation framework for comprehensive analysis of PCD-associated genes.

Results: Our PCDi-based prognostic model outperformed previously published signature models, stratifying cervical cancer patients into two distinct groups with significant differences in overall survival prognosis, tumor immune features, and drug sensitivity. Higher PCDi scores were associated with poorer prognosis. The nomogram survival model integrated PCDi and clinical characteristics, demonstrating higher prognostic prediction performance. Furthermore, our study investigated the immune features of cervical cancer patients and found that those with high PCDi scores had lower infiltrating immune cells, lower potential of T cell dysfunction, and higher potential of T cell exclusion. Patients with high PCDi scores were resistant to classic chemotherapy regimens, including cisplatin, docetaxel, and paclitaxel, but showed sensitivity to the inhibitor SB505124 and Trametinib.

Conclusion: Our findings suggest that PCD-related gene signature could serve as a useful biomarker to reliably predict prognosis and guide treatment decisions in cervical cancer.

Keywords: programmed cell death, prognosis prediction, machine learning, nomogram, cervical cancer

Introduction

Cervical cancer remains the most common gynecological cancer for women worldwide, with 604,000 incidences and 342,000 deaths reported by World Health Organization in its latest report [1]. Unfortunately, developing regions such as Asia and Africa suffer a disproportionate burden of cervical cancer due to limited access to HPV vaccinations and early-cancer screening programs [2]. While surgery is the primary treatment option for cervical cancer, patients diagnosed with locally advanced cervical cancer (LACC) with risk factors usually require adjuvant concurrent chemoradiotherapy (CCRT) after
surgery. However, the heterogeneity of cervical cancer often results in variable clinical outcomes among patients [3].

Recent advancements in bioinformatics have led to the development of numerous prognostic gene signatures for cervical cancer [4–6]. Nonetheless, the application of these prognostic signatures in clinical practice has been hindered by their poor accuracy and unreliability. Therefore, there is an urgent need for a high-quality molecular prognosis prediction model to aid in clinical treatment decision-making.

Programmed cell death (PCD) is a crucial biological process that regulates cell suicide through specific signaling cascades, unlike accident cell death (ACD) caused by external injury or unintentional damage [7, 8]. Various types of PCD mechanisms have been identified, including apoptosis, lysosome-dependent cell death, pyroptosis, immunogenic cell death, necroptosis, ferroptosis, autophagy-dependent cell death, cuproptosis, anoikis, paraptosis, parthanatos, entotic cell death, netotic cell death, alkaliptosis, and oxeiptosis.

Apoptosis is a gene-regulated biological process that enables phagocytes to disassemble and digest injured cells without affecting surrounding cells [9–11]. Paraptosis, mediated by mitogen-activated protein kinases (MAPKs) and inhibited by the multifunctional adaptor protein Alix, is induced by different natural compounds [12, 13]. Lysosome-dependent cell death is triggered by hydrolytic enzymes like cathepsins released into cytosol for cellular component degradations following lysosomal membrane permeabilization [14–16]. Pyroptosis, an inflammatory form of cell death, is activated by the inflammasome multiprotein complex, followed by the release of pro-inflammatory factors [17, 18]. Netotic cell death is caused by the formation of neutrophil extracellular traps (NETs) which are composed of decondensed chromatin and bactericidal proteins [19]. Immunogenic cell death results in an immune response characterized by the secretion of various types of damage-associated molecular patterns (DAMPs) caused by endoplasmic reticulum stress [20, 21]. Necroptosis, another form of inflammatory cell death shares similar morphological features with necrosis but greatly depends on the signaling pathway involving RIPK3 and MLKL [22, 23]. Entotic cell death is caused by a cell invading to an adjacent living cell, leading to the death of the inner cell [24]. Parthanatos is a form of cell death that relies on the activation of poly(ADP-ribose)-polymerase (PARP) and the translocation of mitochondrial-associated apoptosis-inducing factor (AIF), leading to DNA fragmentation and chromatin condensation [25]. Anoikis is a regulated cell death form triggered by the detachment of anchorage-dependent cells from the surrounding extracellular matrix (ECM) due to the loss of cell-matrix interaction [26]. Ferroptosis is initiated by the accumulation of lipid peroxides balanced by the production of reaction oxygen species (ROS) and antioxidant system in an iron-dependent manner [27–29]. Autophagy-dependent cell death is regulated by over 40 autophagy-related genes, driven by autophagic machinery [30]. Alkaliptosis, a pH-dependent cell death form, is induced by intracellular alkalinization through NF-kappaB signaling pathway and carbonic anhydrase 9 (CA9) downregulation [31]. Oxeiptosis is triggered by a high intracellular ROS level, leading to a caspase-independent cell death process through the regulation of KEAP1-PGAM5-AIFM1 pathway [32]. Similar to ferroptosis, the accumulation of the heavy metal copper can induce mitochondrial stress, eventually leading to cell deaths eventually. This process has been termed cuproptosis [33, 34].

Recent advances in PCD studies have led to the development of novel anticancer strategies and drugs that promote cell death, showing promising results in cancer treatment. Specifically, targeting overexpressed proteins like BCL-2/BCL-XL and MCL1 enables the therapeutic induction of apoptosis in tumors [11]. Additionally, a recent study has proposed a novel approach to overcome drug resistance in chemotherapy treatment by inducing ferroptosis [27]. Small molecules targeting caspase-1 can trigger pyroptosis, leading to the destruction of tumor cells in colorectal cancer [35]. Furthermore, sensitizing detached tumor cells to anoikis can prevent tumor metastasis, making it a potential strategy in cancer therapy [36, 37]. Recent studies have reported promising findings have been reported in killing cervical cancer cells by activating PCD mechanisms. Evidence from in vivo experiments have shown the significant impact of Nrf2 on the metastasis of cervical cancer by the enhancement of epithelial-mesenchymal transition (EMT) and anoikis resistance [38]. Moreover, the potential therapeutic agents have been proposed to actively induce PCD activities. It has been reported that necroptosis could be induced in cervical cancer cells by a small anticancer agent called RETRA (REactivation of Transcriptional Reporter Activity) [39]. Besides, as a natural bioflavonoid found in many medicinal herbs/plants, Pinostrobin (PN) exerted anticancer effect to eliminate cervical cancer cells by ROS-dependent apoptosis [40].

Cell homeostasis in multicellular organisms is maintained through the intricate processes of diverse PCDs [8]. However, PCD mechanisms are often impaired in cancer cells, which allows them to resist
and evade various forms of gene-regulated cell death, leading to tumor growth, progression, and metastasis [41]. Our study presents a comprehensive exploration of the intricate relationship between PCD and cervical cancer. By integrating survival-associated genes, we have established a high-quality molecular prognostic model based on a novel indicator known as the programmed cell death index (PCDi), which holds great promise for predicting prognosis and selecting therapeutic regimens for cervical cancer patients. Our study identifies the heterogeneity of cervical cancer based on different PCD mechanisms, contributing to a deeper understanding of the disease and holding significant clinical implications. The development of PCDi highlights the potential for personalized medicine in the management of cervical cancer, ultimately leading to improved patient outcomes.

Materials and Methods

Data collection

PCD processes are regulated by various PCD-related genes. To create an ultimate gene list, we curated regulatory genes for fifteen PCD mechanisms from review articles and databases. Genes related to apoptosis, lysosome-dependent cell death, netotic cell death, entotic cell death, parthanatos, autophagy, alkaliptosis, and oxeiptosis were extracted based on a curated gene list [42]. Additionally, pyroptosis- and immunogenic cell death-related genes were extracted from prior reviews [43, 44]. Gene related to parapthotosis, anoikis, and necroptosis with a relevance score > 0.4 were downloaded from GeneCards database (https://www.genecards.org/) [45]. Furthermore, ferroptosis- and cuproptosis-related genes were downloaded from FerrDb database (http://www.zhounan.org/ferrdb/current/), only driver and suppressor datasets were extracted [46]. The ultimate gene list contains a total of 1,949 PCD-related genes, including 580 apoptosis-related genes, 508 anoikis-related genes, 483 ferroptosis-related genes, 11 cuproptosis-related genes, 209 necroptosis-related genes, 33 pyroptosis-related genes, 34 immunogenic cell death-related genes, 367 autophagy-related genes, 17 parathotosis-related genes, 9 parthanatos-related genes, 15 entotic cell death-related genes, 8 netotic cell death-related genes, 7 alkaliptosis-related genes, 5 oxeiptosis-related genes, and 220 lysosome-dependent cell death-related genes (Supplementary Table S1).

Four datasets from different research institutes were collected for the following analysis. Using Xena platform from USCS Genomics Institute (https://xenabrowser.net/datapages/), we obtained a combined cervical cancer cohort of The Cancer Genome Atlas (TCGA) and The Genotype-Tissue Expression (GTex) samples [47]. To ensure consistency and comparability of clinical data, cancer stage information of patients was standardized according to 2009 The International Federation of Gynecology and Obstetrics (FIGO) Staging Classification (Supplementary Fig. 1). Two tumor samples were excluded because they came from metastatic tumor tissues, and five tumor samples were excluded because of a lack of mRNA expression data (Supplementary Fig. 2A). Molecular and clinical data from CGCI-HTMCP-CC dataset were accessed through National Cancer Institute’s Genome Data Commons (https://gdc.cancer.gov/about-data/publications/CGCI-HTMCP-CC-2020) [48]. Tumor tissues of five patients were excluded because they were ultimately not diagnosed with cervical cancer (Supplementary Fig. 2B). Through Gene Expression Omnibus (GEO) database, microarray and clinical data from GSE52904 and GSE44001 were collected [49, 50]. For high-dimensional data visualization, t-distributed stochastic neighbor embedding (t-SNE) method was employed, using the R package “tsne”.

Identification of the differentially expressed PCD-related genes

To identify differentially expressed genes (DEGs), we utilized samples from TCGA-GTex cervical cancer cohort, consisting of 303 tumor tissues and 13 normal tissues. Raw transcriptome count data was analyzed using the R packages “edgeR”, “limma”, and “DESeq2” [51–53]. A gene identified by at least two algorithms with the selection criteria of adjusted p-value or FDR < 0.05 and |Log2FC| >= 2, was considered as a significant DEG. The absolute value of the logarithm base 2 of the fold change (FC) greater than or equal to 2 is a common criterion for identifying DEGs.

Construction of the PCD-related gene signature by machine learning algorithms

Survival-associated DEGs were screened using univariate Cox regression with a significant threshold of p-value < 0.1. Currently, many types of machine learning were available with unique algorithm design. Without systematic comparison, it was unfair to rely on a particular machine learning algorithm for the model building. We employed ten machine learning algorithms, including survival support vector machine (survivalSVM), elastic network (Enet), least absolute shrinkage and selection operator (LASSO), ridge regression (Ridge), random survival forest (RSF), stepwise Cox regression analysis (stepwiseCox), fitting Cox models by likelihood based boosting (CoxBoost), generalized boosted regression
modeling (GBM), partial least squares regression for Cox (plsRcox), and supervised principal components (SuperPC) based on a 10-fold cross-validation framework to identify the optimal gene signature for cervical cancer (Supplementary Method). A total of 101 combination results were generated. We evaluated the performance of prognostic signatures using a Rank Score, which was calculated using the following formula:

\[
    \text{Rank Score} = \log_{10}(R_{\text{mean}} \times \sqrt{n})
\]

\(R_{\text{mean}}\) stands for the average rank and \(n\) denotes the number of genes in the signature model. Average area under the curve (AUC) of GSE52904 was the mean of 1-year and 2-year AUC values from time-dependent receiver operating characteristic (ROC) analysis. Average AUC of other datasets was the mean of 1-year, 2-year, 3-year, and 5-year AUC values from time-dependent ROC analysis. We chose the optimal signature model with lowest Rank Score, which had a reasonable number of genes and high average rank of Harrell’s concordance index (C-index) and AUC value.

In this study, the prognostic gene signature was established through a combination of two machine learning algorithms stepwiseCox and Enet. Ten genes were identified by stepwiseCox algorithm with backward approach to yield the optimal model with minimum Akaike information criterion (AIC). Subsequently, the regression model was finalized using Enet algorithm with a parameter alpha set to 0.5. For each patient, a PCDi score was calculated using the following formula:

\[
    PCDi = \sum_{j=1}^{n} C_j \times E_j
\]

\(C_j\) stands for the coefficient and \(E_j\) denotes the expression value of each PCD-related gene. To facilitate cross-comparisons of PCDi across the datasets, a linear transformation was employed for data normalization. The normalized PCDi scores had a range from 0 to 1. Based on the median PCDi, cervical cancer patients were separated into PCDi-High and PCDi-Low groups. The R package “stats” was used to perform principal component analysis (PCA) for visualizing high-dimensional data. Kaplan-Meier (KM) analysis was used to investigate the survival difference between two compared groups. The R packages “survival” and “survminer” were implemented for survival analysis and result visualization.

**Functional annotation and enrichment analysis**

The R package “clusterProfiler” was used to perform functional annotation analysis of DEGs based on Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) databases [54]. REVIGO algorithm was implemented to remove redundant GO terms [55]. Gene set enrichment analysis (GSEA) was performed to identify the significant enriched signaling pathways in association with PCDi subgroups.

**Establishment of the nomogram**

The R package “regplot” was used to develop a nomogram survival model for cervical cancer that integrated PCDi and clinical features, including FIGO cancer stage and age at diagnosis. We evaluated the efficacy of the nomogram in prognosis prediction using the R packages “rmda” and “rms” for model calibration and decision curve analysis (DCA). The performance of the nomogram was also assessed by AUC value obtained through time-dependent ROC analysis with the R package “timeROC”.

**Tumor cell infiltration analysis and drug sensitivity prediction**

Infiltrating immune cells in cervical cancer were estimated using TIMER2.0 platform [56]. Tumor Immune Dysfunction and Exclusion (TIDE) algorithm was used to estimate the potential of tumor immune escape and response [57]. The lists of gene signatures of 28 tumor infiltrating lymphocytes and genes encoding for immunomodulators and chemokines were downloaded from TISIDB database (http://cis.ku.hk/TISIDB/index.php) [58]. Single-sample gene set enrichment analysis (ssGSEA) was used to calculate the enrichment scores of 28 tumor infiltrates. The R package “oncoPredict” was utilized to predict drug sensitivities for individual cervical cancer patients [59].

**Formalin-fixed and paraffin-embedded (FFPE) samples acquisition and immunohistochemistry (IHC) staining**

Four FFPE samples were collected from patients diagnosed with cervical cancer from January 2023 to October 2023 at The First People’s Hospital of Foshan. The samples of surgically resected cancerous and normal tissues were preserved after 4% formalin fixation and paraffin embedding treatment. Five µm thick sections were analyzed by IHC staining, as previously described [60]. Briefly, heat-induced antigen retrieval was performed using 10 mM sodium citrate buffer, pH 6.5. Peroxidase activity was quenched with 3% H2O2 and tissues were blocked in 5% bovine serum albumin for 1 hour. Primary MMP1 (1:2000) (10371–2-AP, Proteintech) was added and sections were incubated overnight at 4 °C. Horse-radish peroxidase-conjugated secondary antibody (1:1000) (K8002, Dako Corp., Ltd) and DAB were used.
for detection. Slides were counterstained with hematoxylin.

Statistical analysis and data visualization

R software (http://www.R-project.org) was used for all other statistical analyses, such as Spearman correlation, Kruskal-wallis test, and Mann-Whitney test. Data visualization was carried out using the R package “ggstatsplot”, “ggplot2”, and Sangerbox platform (http://www.sangerbox.com/home.html).

Results

Landscape of PCD-related genes for cervical cancer

In this study, we aimed to develop a reliable prognostic signature for cervical cancer based on the expression of PCD-related genes. Different PCD mechanisms regulate cell elimination that plays a crucial role in cancer development and therapy. PCD-related DEGs are genes that show significant changes in expression levels between cancer and normal tissues, and may reflect the PCD status of cancer cells. We hypothesized that PCD-related DEGs could serve as potential biomarkers for predicting the survival outcomes of cervical cancer patients. To test this hypothesis, we performed the following analyses. We collected 303 and 118 cervical cancer patients from TCGA-CESC and CGCI-HTMCP-CC datasets respectively, as the training cohort. We also collected 55, 173, and 300 patients from GSE52904, TCGA-CESC, and GSE44001 datasets respectively, as the validation cohort. We designed a computational workflow consisting of five major steps to develop the PCD-related prognostic signature for cervical cancer (Figure 1). In the first step, we compiled an ultimate gene list with 1,949 regulatory genes associated with fifteen PCD patterns based on literature and database search. In the second step, we identified PCD-related DEGs from the ultimate gene list by performing differential expression analysis of TCGA-GTEX dataset. In the third step, we further narrowed down the target genes by selecting prognosis-associated genes from the PCD-related DEGs using univariate Cox regression. In the fourth step, we established the most valuable PCD-related prognostic gene signature by applying the combination of 10 machine learning algorithms based on 10-fold cross-validation framework. In the last step, we conducted a group of follow-up analyses to demonstrate the importance of PCD-related prognostic gene signature. More importantly, we used independent datasets to further validate the prognosis prediction of the gene signature.

Using three different algorithms with selection criteria of adjusted p-value or FDR < 0.05 and |Log2FC| >= 2, we identified 2,803 up-regulated genes and 2,380 down-regulated genes in TCGA-CESC cervical cancer dataset (Figure 2A to C, Supplementary Table S2). Among the DEGs, a total of 302 PCD-related genes were identified, including 198 up-regulated and 104 down-regulated genes. A heatmap of Z-score transformed expression levels between tumor and normal tissues is showed in Figure 2E, and two groups are well-separated (Figure 2D). Notably, KEGG pathway analysis revealed that the DEGs were not only implicated in cell death processes such as ferroptosis and necroptosis, but also in the signaling pathways like PI3K-Akt and NF-kappaB pathways, which play a critical role in regulating cell growth and survival (Figure 2F). Furthermore, GO functional annotation analysis showed that the DEGs are involved in the biological processes relating to oxygen level and ROS, which are pivotal mediators of multiple PCD mechanisms (Figure 2G).

PCD-related prognostic gene signature

The overall survival (OS) data from two datasets within the training cohort were analyzed using univariate Cox regression analysis to identify survival-associated genes among 302 PCD-related DEGs. Applying a significant threshold of p-value < 0.1, 80 and 62 genes were identified in TCGA-CESC and CGCI-HTMCP-CC datasets respectively. Twenty-seven genes were consistently identified in both datasets and were subsequently utilized to construct the most robust prognostic gene signature (Supplementary Fig. S3). This procedure was achieved through the combinations of 10 machine learning algorithms, based on 10-fold cross-validation framework. The finalized PCD-related prognostic signature with best performance was established by the combination of stepwiseCox and Enet algorithms (Figure 3, Supplementary Table S3).

The PCD-related prognostic signature consisted of 10 genes that were up-regulated in tumor tissues (Supplementary Fig. S4). Among these genes, three genes were related to anoikis (SPP1, SPIB, FASLG), one gene was related to netotic cell death-related (MMP1), four genes were related to ferroptosis (ALOX15, GLS2, CA9, IFNG), two genes were related to apoptosis (IFNG, FASLG), two genes were related to immunogenic cell death (FOXP3, IFNG), one gene was related to necroptosis (FASLG), one gene was related to lysosome-dependent cell death (CLNK), one gene was related to alkaliposis (CA9 ), and one gene related to autophagy-dependent cell death (IFNG). Notably, FALS, CA9, and IFNG were involved in multiple PCD mechanisms.
Figure 1. Computational workflow for developing the PCD-related prognostic signature for cervical cancer. The workflow includes five major steps. Firstly, PCD-related genes were identified through literature and database search. Secondly, DEGs were identified using TCGA-CESC cervical cancer dataset. Thirdly, univariate Cox regression was performed to identify prognostic genes. Fourthly, machine learning algorithms were used to identify the most valuable prognostic gene signature. Finally, independent validation and follow-up analysis were conducted.
Figure 2. Comprehensive analysis of genes associated to fifteen types of PCD patterns for cervical cancer. (A) Volcano plots displaying the DEGs identified by algorithms DESeq2, edgeR, and limma algorithms. (B) Venn Diagram showing up-regulated DEGs in TCGA-CESC dataset. (C) Venn Diagram showing down-regulated DEGs in TCGA-CESC dataset. (D) t-SNE plot of transcriptomic data of 302 PCD-related DEGs in tumor and normal tissues. (E) Heatmap of PCD-related DEGs between tumor and normal tissues. (F) KEGG pathway annotation analysis based on PCD-related DEGs. (G) GO term annotation analysis based on PCD-related DEGs.
Figure 3. Development of the PCD-related prognostic signature by machine learning algorithms. (A) A total of 101 combinations of machine learning algorithms were used to identify the most valuable prognostic signature for cervical cancer. The C-index and AUC values of each dataset were displayed, and the Rank Score was calculated for each combination based on the rank average and the number of genes in the model. (B) The change of AIC value in stepwise Cox algorithm with a backward approach. (C) The selection of 10 PCD-related genes using Enet algorithm. (D) The selection of lambda.min value to minimize cross-validation error in regression analysis.

We performed KM analysis to assess the association between the expression of genes in the PCD-related prognostic signature and OS. Using a median cutoff, we found that all genes, except for FASLG, IFNG and ALOX15, had significant impact on OS (Log-rank test, \( p < 0.1 \), Supplementary Fig. S5). A novel prognostic indicator, called PCDi was derived from the 10-gene signature model. The PCDi score for each patient was calculated as the sum of the expression levels of 10 genes, weighted by the coefficients derived from Enet regression model. The formula for the calculation of PCDi is shown below.

\[
PCDi = (0.201458 \times \text{SPP1 exp.}) + (0.098878 \times \text{MMP1 exp.}) + (0.130493 \times \text{CA9 exp.}) + (-0.108410 \times \text{ALOX15 exp.}) + (0.304840 \times \text{GLS2 exp.}) + (0.449150 \times \text{FOXP3 exp.}) + (0.198570 \times \text{SPIB exp.}) + (0.781472 \times \text{FASLG exp.}) + (0.675460 \times \text{IFNG exp.}) + (0.988550 \times \text{CLNK exp.})
\]

Based on the median PCDi, we stratified the patients from the TCGA-CESC dataset into two subgroups, PCDi-High (\( n = 152 \)) and PCDi-Low...
Statistical analysis was performed to investigate the associations between PCDi and clinical characteristics in cervical cancer patients (Supplementary Fig. S6A). Out analysis revealed that PCDi was significantly associated with FIGO clinical stage (Supplementary Fig. S6B), tumor size (Supplementary Fig. S6C), tumor metastasis (Supplementary Fig. S6E), and survival status (Supplementary Fig. S6F and G). Notably, even within the same clinical stage, patients from different PCDi subgroup have shown significant OS differences (Supplementary Fig. S7 and S8). However, PCDi was not associated with lymph node invasion in cervical cancer (Supplementary Fig. S6D).

To evaluate the quality of the PCD-related prognostic signature, a total of 33 published mRNA-based prognostic signature for cervical cancer were retrieved through literature search for a comprehensive comparison (Figure 4, Supplementary Table S4). It is noteworthy that the PCD-related prognostic signature exhibited superior performance compared to other published signatures in the prognostic prediction for both OS and disease-free survival (DFS) in cervical cancer patients.

**Prognosis prediction by gene signature in training cohort and validation cohort**

To validate our findings, we compared OS between cervical cancer patients with different PCDi scores in CGCI-HTMCP-CC and GSE52904 datasets. Cervical cancer patients with DFS data were also obtained from TCGA-CESC and GSE44001 datasets for validation.

The PCDi scores were normalized for better comparisons between datasets. Consistent with the results from TCGA-CESC dataset, higher death rate in cervical cancer patients with higher PCDi scores was observed in other datasets (Figure 5A). PCA plot revealed that the cervical cancer patients were well-separated based on PCDi (Figure 5B). The significant OS differences between PCDi-High and PCDi-Low subgroups were observed in both datasets from the training cohort. More importantly, this finding was further validated in the independent cervical cancer datasets, with significant survival differences in both OS and disease-free survival (DFS) in cervical cancer patients.

**PCDi-based nomogram survival model for cervical cancer**

Prognostic factors for cervical cancer were identified using univariate Cox regression analysis. Our finding revealed that patients with high PCDi scores (Hazard Ratio = 2.91, 95% Confidence Interval: 2.16-3.91), as well as those with locally advanced stage (Stage Ib2-IVA, HR = 2.06, 95% CI: 1.17-3.65) or metastasized stage (Stage IVB, HR = 6.12, 95% CI: 2.58-14.51) cervical cancer had higher risk of OS (Figure 6A).

To eliminate false discovery results contributed by the confounding factors, multivariate Cox analysis was performed, which confirmed that PCDi was an independent predictor of OS (HR = 3.19, 95% CI: 2.33-4.39, Figure 6B), highlighting its clinical significance in predicting the prognosis of cervical cancer patients (Supplementary Fig. S9). Based on multivariate Cox regression analysis, a nomogram survival model that integrated clinical features and PCDi was established to estimate 1-, 2-, 3-, and 5-year OS for cervical cancer patients (C-index = 0.789, 95% CI: 0.737-0.842, Figure 6C). High accuracy of the nomogram model in predicting the prognosis of cervical cancer patients was presented in Figure 6E. Additionally, DCA plot showed the nomogram survival model outperformed other predictors in the prediction of survival outcome (Figure 6F). Patients were categorized into Risk-High and Risk-Low subgroups based on the median score given by the nomogram survival model, revealing a significant survival difference (HR = 5.72, 95% CI: 3.18-10.29, Figure 6D). Furthermore, time-dependent ROC analysis demonstrated that the nomogram survival model had better performance than the PCD-related gene signature model in predicting multi-year OS of cervical cancer patients (Figure 6G and H). Our results highlight the clinical significance of PCDi as a prognostic factor and the utility of the nomogram survival model in the prediction of prognosis for cervical cancer.

**Immune features and drug sensitivities for cervical cancer**

To investigate the differences in tumor immune features between two PCDi groups in cervical cancer, we estimated immune infiltrate abundance by calculating the enrichment scores of tumor-infiltrating immune cells using eight different algorithms (TIMER, CIBERSORT, CIBERSORT-ABS, EPIC, ESTIMATE, MCP-COUNTER, QUANTISEQ, and XELL). Additionally, the abundance of 28 tumor immune modulators was quantified using ssGSEA algorithm. Interestingly, our results revealed significant differences in tumor microenvironment (TME) of cervical cancer, with patients having higher PCDi scores exhibiting significantly lower tumor infiltration levels of immune cells (Figure 7A, Supplementary Fig. S10). Furthermore, most immune modulators were negatively correlated with PCDi, demonstrating a similar trend with TMB profiles (Figure 7B).
We utilized GSEA algorithm to further investigate the dysregulated signaling pathways between two PCDi groups. The results showed that immune-related signaling pathways were up-regulated in one PCDi group compared to the other. The integrated analysis of the dysregulated pathways and the clinical outcomes provided new insights into the biological mechanisms underlying cervical cancer progression.

**References**

1. Lin et al. (2023)
2. Liu et al. (2022)
3. Chen et al. (2021)

**Supplementary Information**

- **Supplementary Figure 1**: Heatmap of the dysregulated gene expression in each PCDi group.
- **Supplementary Table 1**: Detailed list of dysregulated genes.

---

https://www.jcancer.org
regulated in PCDi-Low group, such as antigen processing and presentation, B cell receptor signaling pathway, T cell receptor signaling pathway, and so on (Figure 7C). Additionally, we evaluated the potential for tumor immune escape and response using TIDE algorithm, which showed a higher potential of T cell dysfunction was observed in PCDi-Low group (Supplementary Fig. S11A and C), while the prediction scores of T cell exclusion were positively correlated with PCDi (Figure 7D and E, Supplementary Fig. S11B and D).

Furthermore, we analyzed the association between drug’s half maximum inhibitory concentration (IC50) and PCDi to predict the sensitivities of various drugs for cervical cancer (Figure 8A). We observed significant positive correlations between PCDi and IC50 values of commonly used chemotherapy regimens for cervical cancer such as cisplatin, docetaxel, paclitaxel, and gemcitabine (Figure 8B and C), indicating that patients with high PCDi scores were more likely resistant to chemotherapy. Fortunately, we also found that treatment options were available for PCDi-high group, such as SB505124 and Trametinib (Figure 8D and E), proving potential alternative therapeutic options for these patients.

**Experimental validation of MMP1**

For the developed PCD-related prognostic gene signature, we were interested in the contribution of individual genes to the signature. We calculated the relative importance of individual genes of our prognostic gene signature using three machine learning algorithm RSF, GBM, and SuperPC. We found that the relative importance of MMP1 gene to our prognostic gene signature was up 0.216 (Figure 9A). Matrix metalloproteinase-1 is encoded by gene MMP1, also known as interstitial collagenase, involves in the breakdown of interstitial collagens. It has been reported that the MMP1 was up-regulated and affecting lymph node metastasis of cervical cancer through PPAR signaling pathways *in vivo* [61].

---

**Table:**

<table>
<thead>
<tr>
<th>Drug</th>
<th>IC50 (µM)</th>
<th>PCDi</th>
<th>Sensitivity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cisplatin</td>
<td>10</td>
<td>0.25</td>
<td>High</td>
</tr>
<tr>
<td>Docetaxel</td>
<td>50</td>
<td>0.15</td>
<td>Moderate</td>
</tr>
<tr>
<td>Paclitaxel</td>
<td>20</td>
<td>0.30</td>
<td>Low</td>
</tr>
<tr>
<td>Gemcitabine</td>
<td>15</td>
<td>0.20</td>
<td>Low</td>
</tr>
</tbody>
</table>

**Figure 5.** Validation of the PCD-related prognostic gene signature for cervical cancer. (A) Distribution of survival time and status of patients based on increasing normalized PCDi (from left to right) in TCGA-CESC, CGCI-HTMCP-CC, GSE52904, TCGA-CESC (DFS), and GSE44001 (DFS) datasets. (B) Expression levels of 10 PCD-related genes in patients with increasing normalized PCDi. (C) Principle component analysis (PCA) plot of patients based on the expression levels of 10 PCD-related genes. (D) KM estimates of OS in PCDi-High and PCDi-Low group patients.
Figure 6. Development and assessment of the nomogram survival model for cervical cancer. (A) Univariate Cox regression analysis for PCDi and clinical data in TCGA-CESC dataset. The variable with a p-value <0.05 is highlighted in red. (B) Multivariate Cox regression analysis for PCDi and clinical data in TCGA-CESC dataset. The variable with a p-value <0.2 is kept in the regression model. The variable with a p-value <0.05 is highlighted in red. (C) Nomogram survival model developed by integrating PCDi and clinical characteristics. (D) KM estimate of OS in Risk-High and Risk-Low group patients classified by nomogram model. (E) Calibration plot of the nomogram model in predicting multi-year OS. The x-axis and y-axis indicate nomogram-predicted probability and actual probability of OS respectively. The 45-degree line represents ideal prediction, where...
the predicted probability matches the actual probability. The dots represent the prediction sets for different follow-up periods. The vertical line of each dot represents 95% CI of the actual probability. (F) DCA plot evaluating the performance of different predictors in predicting survival status. The x-axis and y-axis indicate the risk threshold and the standardized net benefit, respectively. The curves show the net benefit of using different predictors including Age (red), FIGO Stage (green), PCDi (cyan), and Nomogram (purple) across different threshold probabilities. A predictor with highest net benefit across a range of risk threshold is recommended. (G-H) Time-dependent ROC analysis evaluating the performances of PCD-related gene signature (G) and nomogram model (H) in predicting multi-year OS in TCGA- CESC, CGCI-HTMCP-CC, and GSE52904 datasets.

Figure 7. TME in cervical cancer patients with different PCDi scores. (A) Heatmap displaying the expression profiles of different tumor immune infiltrates between PCDi-High and PCDi-Low groups for cervical cancer. (B) Heatmap displaying the expression profiles of immune modulators between PCDi-High and PCDi-Low groups for cervical cancer.
According to GEPIA2 database (http://gepia2.cancer-pku.cn), the expression of MMP1 was significantly higher in the tumor tissue among different cancer types when comparing to the normal tissue (Supplementary Fig. 12). IHC staining images have shown that the overexpression of MMP1 in tumor tissue compared to normal tissue (Figure 9B). In short, the expression of MMP1 was elevated and may serve as a prognostic marker in cervical cancer.
Figure 9. MMP1 is the most important gene in the PCD-related prognostic signature and is highly expressed in cervical cancer tissues. (A) Relative importance of individual genes to the PCD-related prognostic signature was calculated by three algorithms, including SuperPC (green), RSF (yellow), and GBM (blue). The red bar represents the average relative importance across the three algorithms. MMP1 had the highest average relative importance, indicating that it was the most influential gene in the prognostic signature. (B) Representative IHC staining images for MMP1 in cervical cancer tissues and normal tissues. The expression level of MMP1 was indicated by the intensity of the brown color. Cervical cancer tissues showed higher expression level of MMP1 than normal tissues, suggesting that MMP1 may play a role in the progression and invasion of cervical cancer.

Discussion

To the best of our knowledge, this study represents the first comprehensive analysis of PCD patterns in TCGA-CESC dataset to develop a high-quality prognostic gene signature for cervical cancer. We introduced a novel concept named Rank Score to evaluate the performance of gene signature in prognostic prediction of cervical cancer. The calculation of a Rank Score of a gene signature not only integrated the prediction performance of OS and DFS, but also considered the number of genes included. The optimal signature model with very low Rank Score indicated its high C-index and AUC value in prognosis prediction and a reasonable number of genes. Our analysis resulted in a highly accurate OS prediction model that can aid in therapeutic decision-making, outperforming previously published prognostic signatures and validated by independent datasets. We further established a nomogram survival model that integrates PCDi and clinical features, demonstrating excellent performance in prognosis prediction. The FIGO staging system is one of the most important clinical indicators of the cancers originated from female reproductive system which provides accurate assessment of cancer development.
for appropriate disease management and prognostication in clinical settings. Due to the heterogeneity of cancer, the patients who have the same clinical stage could be further stratified into subgroups with two prognostic outcomes with the help of PCDi (Supplementary Fig. 7). Additionally, our findings highlight significant correlations between PCDi and tumor immune features and drug sensitivities, indicating the potential of PCDi to inform treatment decisions, and have significant implications in clinical practice.

Our PCD-related prognostic gene signature includes 10 genes (SPP1, SPIB, MMP1, ALOX15, GLS2, CA9, IFNG, FOXP3, FASLG, CLNK) that were found to be independent prognostic factors for the OS in cervical cancer. Higher expression of SPP1, MMP1, and CA9 was associated with a poorer prognosis. Secreted Phospho-protein 1 is encoded by gene SPP1, also known as Osteopontin (OPN), which has significant function roles in cancer development, such as cell proliferation and survival [62]. OPN splice variant OPN-c support anchorage-independent growth by inducing the expression of oxidoreductases to avoid anoikis [63, 64]. As a component of NETs, highly expressed MMP1 can promote tumor growth and metastasis in breast cancer cells [65, 66]. CA9 is considered as an endogenous tumor hypoxia marker for cervical cancer, and its overexpression can promote the migration of tumor cells [67, 68]. Lower expression of SPIB, ALOX15, GLS2, IFNG, FOXP3, FASLG, and CLNK was associated with a poorer prognosis. The activation of ETS transcription factor SPIB has been shown to increases anoikis resistance in vitro [69]. In gastric cancer, cancer- associated fibroblasts secreted exo-miR-522 directly targets arachidonate lipoxygenase 15 (ALOX15) to suppress ferroptosis [70]. In hepatocellular carcinoma, glutamine synthases 2 (GLS2) acts as a tumor suppressor by promoting ferroptosis through the production of α-ketoglutarate-dependent lipid ROS [71]. Interferon gamma, encoded by gene IFNG, can induce caspase dependent cell apoptosis in pancreatic cancer cells through the upregulation of procaspase-1 and interferon regulatory factor 1 [72]. Forkhead box P3 (FOXP3), also known as Scurfin, plays a role in immune responses by regulating the development of regulatory T cells [73]. As a member of the tumor necrosis factor family, FAS/FASLG signaling pathways is triggered by the binding of and FASLG (also known as FASL) to induce cell apoptosis. According to in vitro and in vivo evidence, FASLG targeted gene therapy could suppress the tumor growth in head and neck cancer [74]. CLNK, also known as MIST, is an adaptor protein related to SLP76 protein family that regulates multiple immunoreceptor signaling pathways in a LAT-(linker for activation T cells) dependent manner [75]. Among the 10 genes, we found the highest relative importance of MMP1 in the prognostic model. Our IHC results of cervical cancer patients further validated the overexpression of MMP1 in tumor tissue compared to normal tissue.

Tumor microenvironment is a complex interplay of signaling molecules, structural elements such as extracellular matrix, and various types of cells including stromal cells, immune cells, and tumor cells [76]. The dynamic interactions between these components have profound impacts on cell survival, tumor growth, local invasion, and metastasis [77]. Immune cells, as a critical component of TME, have dichotomous functions of either suppressing tumor formation or promoting tumorigenesis [78]. Tumor cells are under surveillance by the immune system and can be attacked by various immune cells, such as cytotoxic T cells (CD8+), which play a crucial role in killing tumor cells by recognizing tumor antigens and subsequently suppressing tumor growth [77, 78]. Infiltrating B cells, on the other hand, are involved in antigen production, antigen presentation, and secretion of cytokines instead of directly targeting tumor cells [77]. In this study, we found that the enrichment of B cells and CD8+ T cells was negatively correlated with PCDi in cervical cancer patient, indicating lower infiltration of these immune cells in PCDi-High group. A review study has reported that higher infiltration of immune cells, including B cells and CD8+ T cells, is associated with better prognosis [79]. Consistent with previous findings, a worse prognosis in cervical cancer patients with higher PCDi scores was observed in this study. Additionally, TIDE analysis revealed higher potentials of T cell dysfunction and T cell exclusion in PCDi-High group, supporting the significant difference in tumor immune features between two PCDi subgroups. Furthermore, we investigated the sensitivities of various drugs for cervical cancer by analyzing the association between PCDi and IC50. Higher IC50 value of commonly used chemotherapy regimens such as cisplatin, docetaxel, paclitaxel, and gemcitabine, were found in PCDi-High group, indicating high chemotherapy resistance in cervical cancer patients with high PCDi scores, which could potentially explain their poor prognosis. However, there are still treatment options available for patients in the PCDi-High group who have high sensitivity to these drugs, such as SB505124 and Trametinib. SB505124 is a selective inhibitor which targets transformation growth factor beta type I (TGF-beta) receptors ALK4, ALK5, and ALK7 and inhibits downstream Smad signaling [80]. Previous studies
have shown that SB505124 can restrain the migration and invasion of breast cancer cells [81]. Trametinib, an FDA approved mitogen-activated protein kinase kinase (MEK) inhibitor, is used to treat melanoma patients with BRAF V600E. A phase II clinical trial is currently underway to evaluate the efficacy of combining Trametinib and AKT inhibitor GSK21411795 for the treatment of recurrent cervical cancer with PIK3CA and KRAS mutation [82].

In this study, we observed excellent performance of the PCD-related gene signature and the nomogram survival model in both training and validation cohorts. However, there are two limitations that should be acknowledged. Firstly, all tumor samples analyzed were retrospectively recruited, additional datasets with larger sample sizes, high quality, and longer follow-up periods will be required for further validation. Secondly, while our study highlights the importance of PCD-related genes as prognostic markers, there is still insufficient knowledge about some of these genes. Further research, particularly in vivo experiment, is warranted to better understand their roles in cervical cancer.

Conclusions

In conclusion, we proposed a novel prognostic gene signature related to PCD for cervical cancer patients using a machine learning-based framework that can stratify them into two groups with significant differences in prognosis, tumor immune features, and drug sensitivity. Our signature model outperformed than previously published gene signatures in predicting patient’s prognosis and demonstrated robustness and reproducibility in both training and validation cohorts. This signature model could serve as a valuable biomarker for identifying high-risk patients who may require more intensive treatment options or frequent disease surveillance. Moreover, we developed a nomogram survival model that integrates PCD-related gene signature and clinical characteristics to enhance the clinical utility of our signature model. Overall, our study highlights the potential usefulness of PCD-associated genes in predicting cervical cancer prognosis and emphasizes the importance of personalized treatment approaches tailored to individual patient characteristics. Our findings could have important implications in improving the management of cervical cancer patients.

Supplementary Material

Supplementary methods, figures and table legends. https://www.jcancer.org/v15p1378s1.pdf
Supplementary tables. https://www.jcancer.org/v15p1378s2.xlsx

Acknowledgements

Funding

This work was supported by National Natural Science Foundation of China [grant number: 81302281] and Medical Scientific Research Foundation of Guangdong Province, China [grant number: A2016176].

Author contributions

All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by Wei Wang, Pengchen Chen, and Songhua Yuan. The first draft of the manuscript was written by Wei Wang and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Data availability

The datasets TCGA-GTEx/TCGA-CESC for this study can be found in the Xena platform from USCS Genomics Institute [https://xenabrowser.net/]. The datasets CGCI-HTMCP-CC for this study can be found in the National Cancer Institute’s Genome Data Commons [https://gdc.cancer.gov/]. The datasets GSE52904 and GSE44001 for this study can be found in the Gene Expression Omnibus [https://www.ncbi.nlm.nih.gov/geo/].

Ethics approval

This retrospective study was approved by the institutional review board and ethics committee of The First People’s Hospital of Foshan.

Competing Interests

The authors have declared that no competing interest exists.

References


